Ahmed Hossain, Ph.D.
Associate Professor,
Health Services Administration
University of Sharjah.

Biostatistics- 0504 252
Tutorial 1: Measurement

Biostatistics

Introduction

- Public health is primarily concerned with disease prevention in human populations, and epidemiology is the branch of public health that attempts to discover the causes of disease to make disease prevention possible. Public health investigations use quantitative methods, which combine the two disciplines of **epidemiology** and **biostatistics**.

Statistics It is a science concerned with:

- Collection of data
- Presentation of the collected data
- Analysis and interpretation of the results.
- Making decisions based on such analysis

Biostatistics Biostatistics is that branch of statistics that deals primarily with the biological sciences and medical/health-related disciplines.

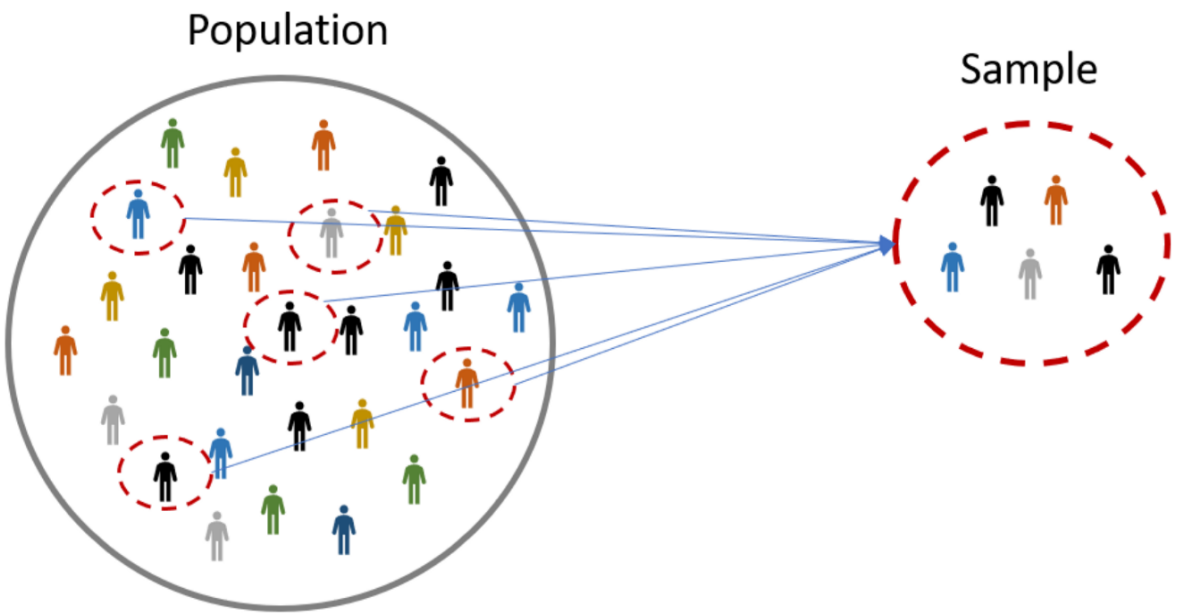
Population and Sample

Population A population is a collection of persons/things or characteristics in which we have the interest to investigate.
For Example, the collection of persons living in Dubai city who test positive for hepatitis C.

Sample A sample is a subset of a population.
Example, we may refer to a sample of 50 men over age 65 who suffer from hypertension or to a sample of 50 blood pressures taken on 50 men over age 65 who suffer from hypertension.

Population and Sample

A sample should be a **representative part** *of the population.*



Variable

Variable A characteristic that takes on more than one value is termed a variable. A variable may also be called a data item.

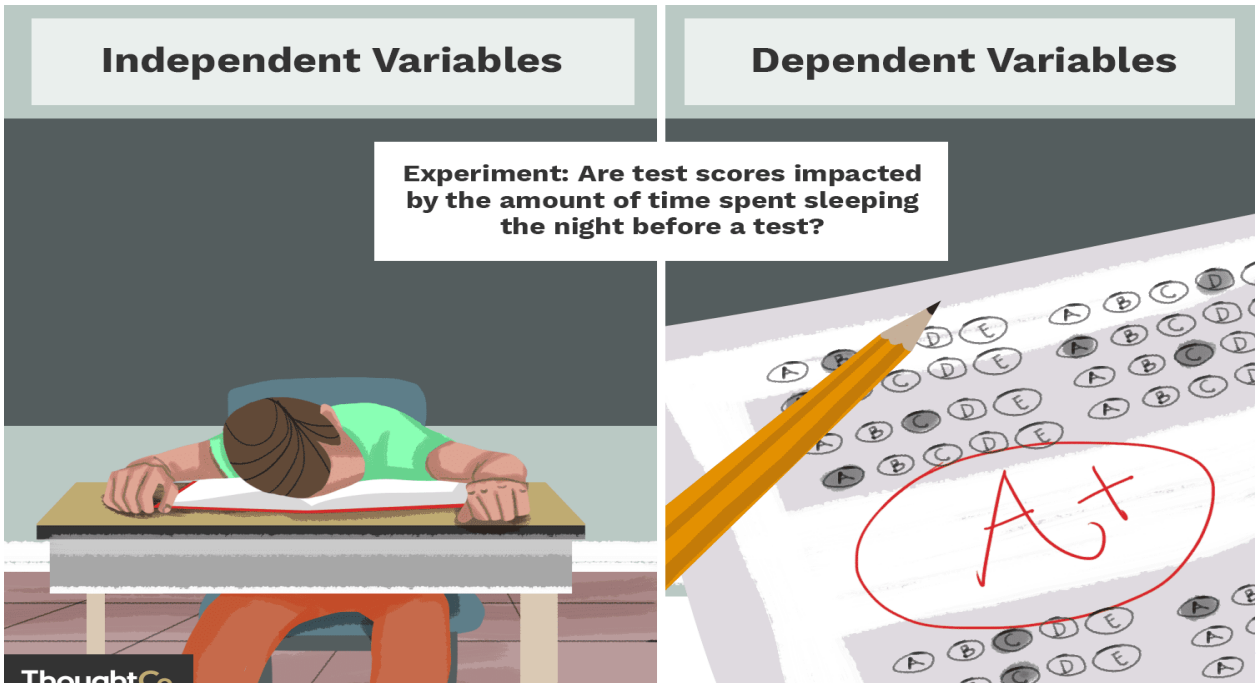
Examples of variables Age, gender, business income and expenses, country of birth, capital expenditure, class grades, eye color, and vehicle type are examples of variables. If a sample consists of 50 males, then gender is not a variable in this sample but is termed a constant. If the sample is made up of males and females, then gender is a variable in this sample.

Biostatistics

Independent and dependent variables

Independent (explanatory) variable is the cause. Its value is independent of other variables in your study.

Dependent (Response) variable is the effect. Its value depends on changes in the independent variable.



Exercise

Find out Independent and dependent variables?

Example 1 Evidence is lacking on the impact of **smoking** on **colorectal cancer (CRC) risk** (overall and by **age** at diagnosis) by polygenic risk score (PRS) levels, and it is unclear how the magnitude of CRC risk associated with smoking compares to the magnitude of genetically determined risk.

Example 2 **HBV infection** was more concentrated among population with high **economic status**.

Example 3 Tens of millions of Americans have **a chronic health condition** that increases their risk of **severe illness from COVID-19**.

Example 4 **SARS-CoV-2** is not just a respiratory virus that affects **the lungs**. It can also affect the **stomach, intestines, heart, blood vessels, liver, kidneys, and the immune system**

A questionnaire

A SNAPSHOT OF A DATA COLLECTION FORM

Module 1: Demographic information of the respondent									
ID	**Name** (First 3 letter)	**1.1 Category of respondent**	**1.2 Sex**	**1.3 Age**	**1.4 Marital status**	**1.5 Highest level of school attended**	**1.6 Primary occupation**	**1.7 Pregnancy**	**1.8 Lactation**
		1. HH Head 2. Others	1. Male 2. Female 3. Third gender	In years	1. Married 2. Unmarried 3. Widowed 4. Divorced/ Separated	1. Primary (1-5 years) 2. Secondary high school (6-10 years) 3. Higher secondary (11-12 years) 4. University or higher (>12 years) 5. Madrasa 6. No schooling	1. Agriculture 2. Livestock 3. Fisheries 4. Service holder 5. Business 6. Day laborer 7. Professional (Physician/lawyer/teacher) 8. Productive work at HH 9. Driver 10. Student 11. Housewife 12. Beggar 13. Unemployed 14. Others	Ask the female respondents aged less than 45 years - If she was ever pregnant in last one year or currently pregnant for at least 3 months 1. Yes 2. No	Ask the female respondents aged less than 45 years - If she was breast feeding her child/children any time during the last one year 1. Yes 2. No
01									
02									
03									
04									
05									

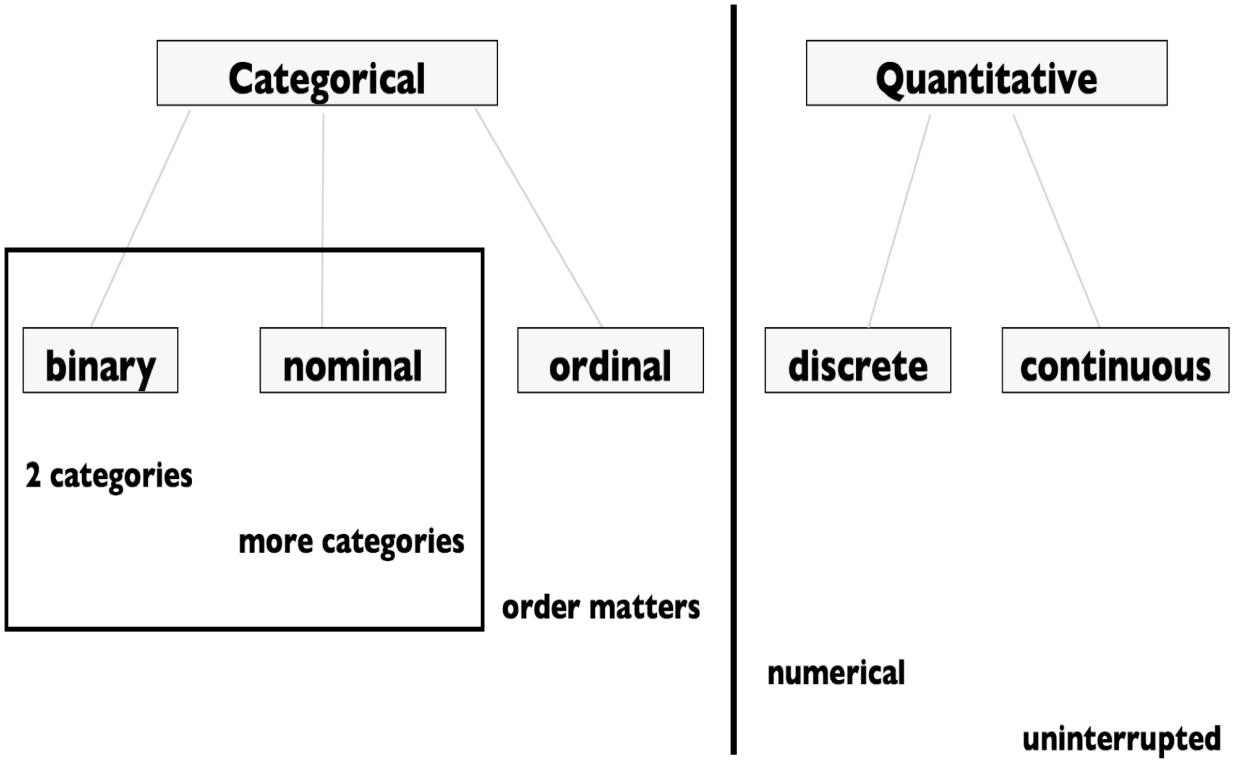
A SNAPSHOT OF A DATA TABLE

	AGE	SEX	HIV	ONSET	INFECT
Obs01	24	M	Y	12-OCT-07	Y
Obs02	14	M	N	30-MAY-05	Y
Obs03	32	F	N	11-NOV-06	N

- Each **row** corresponds to an **observation**
- Each **column** contains information on a **variable**
- Each **cell** in the table contains a **value**

Biostatistics

Data Types/ Measurement scales



Biostatistics

Data Types/ Measurement scales

Nominal It measures values that fall into categories with no natural numerical value. Nominal data is often coded numerically, but the codes are just alternate names. For Example, 0 for females and 1 for males.

Ordinal falls into categories that can be qualitatively ordered but have no intrinsic numerical value. Ordinal data can be 'ranked.' E.g., education.

Discrete data measured quantities that take on specific values, usually integers. E.g., the number of traffic accidents, infant deaths, etc.

Continuous data measured quantities not restricted to specific values. E.g., birth weight, blood pressure, etc.

Exercise

Find out nominal, ordinal or quantitative variables?

Example 1 Age

Example 2 Age groups

Example 3 Sleeping hours

Example 4 blood pressure (SBP and DBP)

Example 5 Presence of hypertension.

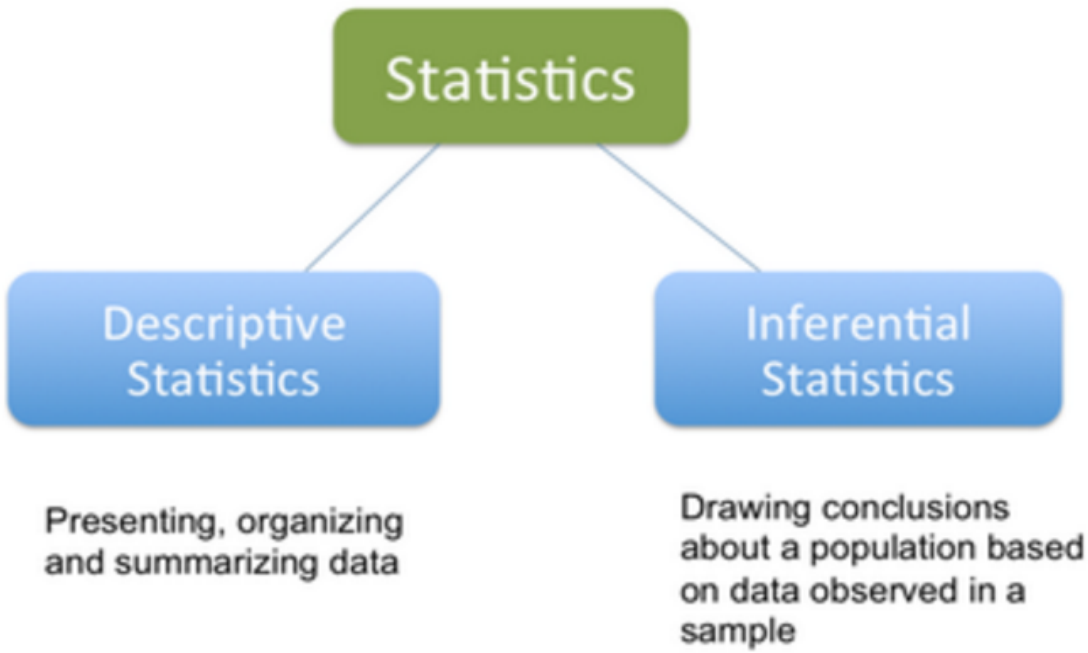
Example 6 Blood type.

Biostatistics

Central dogma of statistics

Parameter Any summarization of the elements of a population. E.g., the average of the blood pressures that make up a population.

Statistic Any summarization of the elements of a sample. E.g., the average of the blood pressures that make up a sample.



Descriptive statistics: Exploratory Data Analysis

- ALWAYS look at your data!
- If you can't see it, don't believe it!
- EDA allows us to:
  1. Visualize distributions and relationships
  2. Detect errors
  3. Assess assumptions for confirmatory analysis
- EDA is the first step of data analysis